

FoodNet Population Survey

Existing Collection In Use Without an OMB Control Number

Supporting Statement - Section B

April 11, 2016

Program Official/Project Officer

Olga Henao, PhD
Division of Foodborne, Waterborne, and Environmental Diseases,
National Center for Emerging and Zoonotic Infectious Diseases,
Centers for Disease Control and Prevention
1600 Clifton Road NE MS C-09, Atlanta, GA30333
Office: (404) 639-3933
Email: dot8@cdc.gov



- Goals of study
 - a. Provide necessary data to estimate the burden of unreported and reported acute diarrheal illness in the catchment area, for use in modeling such burden in the United States more broadly.
 - b. Assess the frequency of important exposures commonly associated with foodborne illnesses in the catchment area, for use in modeling such burden in the United States more broadly.
- Intended use of resulting data:
 - a. Populating models designed to estimate prevalence of diarrheal illnesses in the United States
 - Assess efficacy of interventions implemented to reduce diarrheal diseases
 - Develop metrics for diarrheal disease food safety policies and programs
 - Estimate costs of diarrheal illnesses
 - Inform industry, academic and public health diarrheal disease research and food safety activities
 - b. To provide estimates of food consumption and other exposures in the catchment area for use to guide generation of hypotheses during outbreak investigations
 - c. To provide estimates of food consumption and other exposures in the catchment area to assist in the study of factors (e.g. specific foods, routes of exposure, settings, healthcare seeking behavior) associated with sporadic illness
 - d. To provide data to estimate changes in healthcare seeking behavior and diagnostic testing practices (stool testing) for diarrheal illnesses in the catchment area

3
7
9
11
13
13
of

- Methods for data collection — Multi-mode survey (landline, cell phone, web, mail)
- Subpopulation to be studied — Residents of the FoodNet catchment area
- How data will be analyzed — Descriptive statistics and regression modeling

Data Collection Procedures

1. Respondent Universe and Sampling Methods

The FoodNet catchment area is determined as part of a competitive grant process and has been consistent since 2004. The catchment area includes 15% of the United States population (48 million persons) and consists of Connecticut, Georgia, Minnesota, Maryland, New Mexico, Oregon, Tennessee and selected counties in California (Alameda, San Francisco, and Contra Costa), Colorado (Adams, Arapahoe, Denver, Douglas, Jefferson, Boulder, and Broomfield), and New York (Albany, Allegany, Cattaraugus, Chautauqua, Chemung, Clinton, Columbia, Delaware, Erie, Essex, Franklin, Fulton, Genesee, Greene, Hamilton, Livingston, Monroe, Montgomery, Niagara, Ontario, Orleans, Otsego, Rensselaer, Saratoga, Schenectady, Schoharie, Schuyler, Seneca, Steuben, Warren, Washington, Wayne, Wyoming, and Yates.) The respondent universe for the Population Survey is comprised of residents who live in the FoodNet catchment area.

There are advantages of conducting the Population Survey in the same geographic areas served by FoodNet:

1. FoodNet conducts active surveillance for 9 enteric pathogens, including collection of demographic, geographic, testing practices, and outcome data. The ability to pair the Population Survey data with these active case-based surveillance data is unique to FoodNet catchment areas.
2. By surveying the population that gave rise to the foodborne illnesses captured in FoodNet we have greater confidence that our pathogen-specific adjustments for underdiagnosis of foodborne illness (resulting from medical care seeking and specimen submission) are appropriate.
3. Through routine, active surveillance, FoodNet collects standardized information on food and water consumption and environmental exposures from persons with *Salmonella* and *Campylobacter*. Collecting this same information through the Population Survey from the general population in these FoodNet sites gives us a readily available comparison group which will allow us to identify potential risk factors to inform attribution estimates and as hypothesis to be tested in future research studies.

FoodNet sites were not selected to be representative of the US population. Although assessments of the FoodNet population suggest that, other than underrepresentation of Hispanics, the demographic characteristics of FoodNet catchment area do not differ substantially with US Census,⁶ it is important to note that such a comparison masks regional differences in relationships among demographic factors, such as income, education, and race/ethnicity with urban/rural factors. When using FoodNet data to populate national models, care must be taken not only to adjust foodborne illness estimates to account for lack of geographic coverage, but also to maximize the transparency of the limitations of the dataset when communicating results. Acknowledging the limitation on representativeness, we believe that by the unique benefits offered by pairing the Population Survey with active surveillance of FoodNet generate valuable data.

As collecting data from the entire population in the FoodNet catchment area is not feasible, a sampling strategy will be employed. A total of 36,000 completed questionnaires will be administered over a two year period (150 completed questionnaires per month per site for 24

months). This number was determined to be an adequate sample size given the expected prevalence of diarrhea in the community.

We have contracted ICF International (ICF) to implement our survey.¹ ICF has experience supporting the CDC, and expertise drawn from 25 years of continuous data collection for the Behavioral Risk Factor Surveillance System (BRFSS), and three previous cycles of the FoodNet Population Survey.^{1,2} ICF has experience in designing and conducting dozens of multi-mode surveys in a variety of languages for Federal, State, and local governments. To maximize the populations reached, a multi-frame, multi-mode (mail, web and phone) design described below will be used.

Multi-Frame, Multi-Mode Design

One of the objectives of our design is to determine the combination of modes that will maximize the population reached, we are planning to use initial 50/50 split of completed interviews from the two samples: 18,000 from the address-based sample (ABS) sample and 18,000 from the dual-frame random digit dialing (RDD) sample. We will use an optimal allocation for allocating the RDD sample to cell phone and landline; this allocation minimizes the cost-per-interview and maximizes the final sample’s efficiency. Surveys will be conducted in Spanish and English. A six-stage translation process will be used including: forward translation by two independent and certified bilingual translators, synthesis of forward translations by a third independent and certified bilingual translator, back translation by a fourth independent translator, review by an expert committee, pre-testing, and submission and appraisal. Optimal allocation by site is illustrated in the table below.

Allocation of Completed Interviews to the ABS and Dual-frame RDD conducted in English and Spanish				
		ABS	Cell	Landline
California	Alameda, Contra Costa, San Francisco	1800	835	965
Colorado	Adams, Arapahoe, Boulder, Broomfield, Denver, Douglas, and Jefferson	1800	1042	758
New York	Albany, Allegany, Cattaraugus, Chautauqua, Chemung, Clinton, Columbia, Delaware, Erie, Essex, Franklin, Fulton, Genesee, Greene, Hamilton, Livingston, Monroe, Montgomery, Niagara, Ontario, Orleans, Otsego, Rensselaer, Saratoga, Schenectady, Schoharie, Schuyler, Seneca, Steuben, Warren, Washington, Wayne, Wyoming, and Yates	1800	599	1201
Connecticut	All	1800	549	1251
Georgia	All	1800	979	821
Maryland	All	1800	677	1123
Minnesota	All	1800	900	900
New Mexico	All	1800	945	855
Oregon	All	1800	947	853
Tennessee	All	1800	958	842

Sampling Frame and Respondent Screening

RDD Landline and Cell Phone Frames

The RDD frame originates from the North American Numbering Plan Administration, which governs the assignment of 1,000-blocks to service providers. A 1,000-block is the series of 1,000 telephone numbers defined by the last three digits of a 10-digit phone number (NPA-NXX-Z000 - NPA-NXX-Z999). The 1,000-blocks dedicated to cell service or landline service are identified by codes from the Telcordia® LERG (Local Exchange Routing Guide). Those dedicated to landline service comprise the landline frame, while those dedicated to cell service comprise the cell phone frame.

Landline

Telephone lines are not linked to a physical location, but are generally associated with particular geographic areas. We will identify geographic associations for each of the 10 geographic areas. For landlines, each 1,000-block of telephone numbers is associated with a single geographic area by tallying the number of geocoded landline households in each geographic area. The 1,000-block is assigned to the geographic area with the most number of geocoded telephones (the rule of plurality).

We will select the landline sample using RDD with equal probabilities of selection (EPSEM) from working banks. A “working” bank is a 100-block (NPA-NXX-ZZ00 - NPA-NXX-ZZ99) where at least one telephone number is assigned to residential service. This frame definition is improved over traditional list-assisted frame, in which blocks with one or more “listed” telephone numbers were include in the frame. The traditional list-assisted frame excluded zero-blocks, which typically excluded about 5% of residential households. This frame definition is consistent with CDC’s national BRFSS landline sample.² All listed and unlisted numbers in working landline telephone banks are eligible for selection in the sample. The sampled telephone numbers are purged for known businesses.

For landlines, telephone numbers are classified as high-density or low-density based on whether the telephone number is listed in telephone directories or unlisted. Listed telephone numbers are more likely to be working residential numbers, therefore the high-density stratum will be highly concentrated with eligible respondents. On the contrary, the low-density stratum will include households that have unlisted numbers as well as many non-working numbers, resulting in a low concentration of eligible respondents. We will select a disproportionate sample by proportionally oversampling the high-density numbers to increase the efficiency of reaching residential households. We do this with a two-phase sample approach called double sampling for stratification. We select a RDD sample as described above and match the selected numbers to telephone directories to classify it as listed (high-density) or unlisted (low-density). We then select all listed telephone numbers and a subsample of unlisted numbers such that the proportion of listed sample is 1.5:1 relative to unlisted sample.

Cell Phones

For cell phones, we will identify the “rate centers” associated with each geographic area. A rate center is the midpoint of the rate area (generally a town, county or a group of towns or counties) covered by a telephone bank (exchange or 1,000-block) of numbers. The rate center represents the geographic location where the cell number is originally assigned. While cell

phones are portable to other geographic locations, the location of the rate center is an indicator of the location of the cell phone.

We will select the cell phone sample using RDD with equal probabilities of selection (EPSEM). All telephone numbers from the cell frame will be manually dialed in accordance with laws prohibiting cell numbers from being called by an automated dialer. We will use a service provided by Marketing Systems Group (MSG) called CellWINS, a screening process to identify and remove inactive cell phone numbers. Eliminating “inactive” cell phone numbers reduces the amount of labor spent manually dialing and calling non-working cell phone numbers. Our internal research suggests that the CellWINS pre-screening process will reduce the cost-per-interview by 8–10%.

For the cell phone sample in California, Colorado, and New York, we will use double sampling for stratification to increase the efficiency of reaching respondents residing in the sampled geographic areas. To do this, we will select a sample from the identified rate centers for each area. Then, we match the cell phone sample to external databases to obtain the ZIP code for the billing address for the cell phone service. We then stratify the sample based on whether the billing ZIP code matches to a ZIP code located in the target geography (match-in), matches to a ZIP code located outside the target geography (match-out), or there is no match for the cell phone number (unmatched). We then select a proportionate sample with an oversample on the match-in stratum relative to the unmatched and match-out strata. Sampling from the unmatched stratum is necessary since a large percentage of cell phone numbers will not match to a ZIP code. While a number of these cell phone numbers will not be working, there are some numbers that are active cell phones. Therefore, we include these numbers in the sample to maximize coverage of the geographic area. Based on past experience, we expect about 50% of the cell phone numbers to match to a ZIP code.

Our strategy on the match-out cases will be adaptive. For the first quarter, we will monitor the percentage of match-out cases that actually live in the geographic area (this can happen if the billing ZIP is not the respondent’s home address). If the percentage is small and the match-out stratum represents less than 10% of the population (i.e. match-in and unmatched provides 90% coverage), we will eliminate further sampling in this geographic area.

Respondent Screening

Since we will reach households where more than one person may be survey-eligible, we will screen the household to determine eligibility and then randomly select one person to participate. The selection will be based on a probability or quasi-probability (e.g. next/last birthday) selection.

ABS

The source of the ABS frame will be the Computerized Delivery Sequence File (CDSF), a list of addresses that originates from the USPS. Covering 98% of households, the CDSF provides a comprehensive frame that will reach the entire population living at addresses that receive mail delivery.

One of the benefits of ABS is that addresses are linked to a physical location. Each address on the CDS is geocoded (latitude and longitude) to a physical location. Addresses that fall within one of the 10 geographic areas will be eligible for selection.

We will design and select the sample using Virtual Genesys, which licensed by ISF from MSG. The ability to select samples in-house provides us with the flexibility to design efficient sampling frames, select stratified random samples, and adapt the sampling design to the specific needs of the FoodNet Population Survey. We will include all residential addresses including city-style addresses, P.O. boxes, central drop points, and rural-route addresses. To maximize coverage of the population, our sampling frame will include units identified by the USPS as seasonal and vacant.

Determining the Within-Household Respondent Selection

A number of respondent selection methods have been tested for ABS mail survey including in the BRFSS. HINTS adopted the next birthday method for within-household respondent selection for its survey cycle. Similarly, Messer and Dillman used the next-birthday method for within-household selection in their ABS multi-mode survey of Washington State. Hence, the birthday method is becoming the standard approach to within-household selection for self-administered surveys as it has become for RDD surveys. However, given that next/last birthday methods tend to result in a higher completion rate by the person who opens the mail, this may not be optimal for a within-household selection that includes children. We have developed a Kish grid selection method based on pre-printed randomization (different for every survey) to mitigate this selection bias. This selection methodology will be tested in comparison to next/last birthday methods. Similarly for telephone respondent selection, we are currently testing the use of last/next birthday with the traditional BRFSS selection methodology.

Obtaining Parental Consent for Children Ages 12-17

Obtaining parental consent for children ages 12–17 presents challenges for all survey modes. For the landline survey, we will follow the process used in previous FoodNet Population Surveys by rostering the household and randomly selecting one person; if the selected respondent is between the ages of 12 and 17, the interviewer will require verbal parental consent before conducting the interview with the minor. For the web survey, respondents will be asked their age at the beginning of the survey; if under 18, parental consent will be required on the web form before the respondent can continue. For the mail survey, a consent form will be included as part of the mail packet. Gaining parental consent is particularly challenging when calling cell phones. Cell phones are considered personal devices and are portable. There is a high likelihood that the person answering the telephone will not be in the presence of a selected minor. Similarly, if a minor is reached (12–15% of cell phones), there is a high likelihood that a parent is not available. To simplify the selection, we will develop a child selection procedure based on calling a parent cell phone first to obtain permission to speak with the teen. We will then make an arrangement to speak with the child.

2. Procedures for the Collection of Information

An integrated data collection platform that has been used in numerous surveys and can generate screens adapted for each data collection mode will be used by ICF.¹ This approach offers several benefits including standardized skip patterns and logic rules, quotas, and databases across multiple modes; more efficient, accurate tracking and reporting across all survey modes; and complete flexibility for respondents (they can complete part of the survey by phone or web, return to the survey at a later time, and seamlessly pick up where they left off).

Determining Calling Protocol/Contact Strategy

We plan to follow the BRFSS protocol for the landline and cell phone RDD to maintain continuity with the previous survey designs. The 15-attempt landline and eight-attempt cell protocol over a 30-day period is an extended calling protocol which permits higher contact rates and greater opportunities to complete interviews and convert refusals than shorter protocols. The median response rate for dualframe BRFSS surveys is 44%.

For the ABS sample, we plan to conduct a web/mail dual-mode survey. We will initially make multiple attempts to complete via the web. Then, we will provide an option to complete a mail survey at the last contact attempt. We do not propose a telephone follow-up for the ABS sample because only about a quarter of the sample match to a phone number, and these matches will be limited to landline phones importantly, the characteristics of this subsample of persons with listed landline telephone numbers is very different from the rest of the sample. For instance, using BRFSS data from 10 states, ICF found that unlisted households had a higher percentage of Hispanic respondents (4.8% listed, 9.3% unlisted); black respondents (5.7% listed, 10.4% unlisted); adults ages 18–44 (12.8% listed; 23.4% unlisted); and renters (18.3% listed, 32.1% unlisted). By increasing the proportion of completes from this subsample, the selective follow-up efforts will increase the bias in the estimates as it modestly increases the response rate.

Our proposed contact strategy is illustrated in the box below. By emphasizing the typically lower cost web-based administration in Phase 1, this multi-mode study design will offer a cost-effective data collection while maximizing the number of responses via computer-assisted self-interview; this will provide cleaner data than surveys completed by mail. Our design includes use of a \$5 pre-incentive, which has been shown to increase response rates by as much as 20 points, the use of a web card with instructions for accessing the web survey (ideally reaching less experienced web users), and multiple reminders—all features that have been shown to significantly increase response in ABS “push to web” surveys. Once “phase capacity” has been achieved, the mail survey as a final contact allows response from the entire population (including non-internet users) to minimize coverage bias. Thus, we use a sequential multi-mode approach (i.e. only one mode option at a time) to maximize response. Concurrent mode choice (i.e. offering multiple modes at once) has been repeatedly found to decrease response rates.

Phase 1

Contact 1: Introductory letter that outlines the purpose of the research. Includes:

- URL for completing the survey on the web
- Card with detailed, graphical instructions for accessing the web survey

Contact 2: Thank you letter to those who have completed/reminder to those who have not (two weeks after start). Includes:

- URL for completing the survey on the web
- Card with detailed, graphical instructions for accessing the web survey

Contact 3: Thank you letter to those who have completed/reminder to those who have not (four weeks after start). Includes:

- URL for completing the survey on the web
- Card with detailed, graphical instructions for accessing the web survey

Phase 2

Contact 4: Survey packet (six weeks after start). Includes:

- Cover letter that outlines the purpose of the research (no URL included)

- Full-color mail survey
- Business reply envelope (BRE)

True to responsive design, we will use the first quarter of data collection to evaluate the design features' cost benefits. For instance, literature consistently suggests that a prepaid incentive is beneficial to response rates. However, is a \$2 or \$5 incentive more cost beneficial when balancing the response rate and overall costs? We propose split-sample experiments, and after evaluating, we will conduct the remaining data collection with the most cost-effective strategy, diverting resources to other strategies to increase response. Similarly, we have planned for three invitations to the web survey. We will evaluate whether phase capacity occurs after two invitations. If so, the third contact could be eliminated and those resources used elsewhere. During research planning, we will work with CDC to develop the responsive survey design framework in order to design an effective, sustainable long-term data collection strategy.

Respondents will be given the name of a person at CDC to contact if they have any questions. No websites will be used for data collection in this study.

The survey instrument consists of a section for participant screening and consent or assent, 8 main sections, and a closing statement. The section topics are as follows: Sections 1 and 2: Food exposures, section 3: food contact practices and beliefs, section 4: animal contact, section 5: drinking and recreational water, section 6: travel, section 7: health, section 8: community. All questions are multiple choice and some have a text box for further explanation if the choice is 'Other'. The interview sample will be 'split' which means that no one will answer all questions. At the start of the survey respondents are randomly assigned to splits A or B. Respondents assigned to split A, complete sections 1, 3, 5, 6, 7, and 8. Respondents assigned to split B will complete sections 2, 4, 5, 6, 7, and 8. The survey is designed to take 20 minutes. Prior to launch of the survey, the questionnaire will be pre-tested with up to 200 different sites, including cognitive-, usability-, and pilot-testing.

Data will be stored in databases by the contracting company. The contractor will provide CDC with two datasets quarterly (one containing all records and one containing completed interviews), and a final technical report that includes methodology and response rates.

A copy of the pre-notification letter, telephone script, and questionnaire is included in Attachment D.

3. Methods to Maximize Response Rates Deal with Nonresponse

Concerns about low response rates coupled with growing challenges related to traditional telephone surveys and the emergence of web-based data collection technologies have prompted changes in approach and data collection mode for several large national populations surveys such as the Health Information National Trends survey, the National Household Education Survey, and the ICF-conducted Motor Vehicle Occupant Safety Survey. Thus, to maximize the populations reached, a multi-frame, multi-mode (mail, web and phone) design will be used. A multi-frame sampling design can potentially increase the proportion of the population that can be effectively reached. Address based sampling (ABS) and dual-frame random digit dialing (RDD) each cover 97–98% of the population—despite this near complete coverage for either frame, it is unclear what proportion of the population can actually be reached. For example, while 98% of households are reachable through a dual-frame RDD frame,

it is not clear what proportion of this population is likely to answer their telephone, particularly if the caller is unknown. Similarly, ABS covers virtually all households, but mail without names may be effectively undeliverable for households at delivery points, where mail is delivered for multiple people at a single site. Moreover, issues of functional literacy may reduce the effective coverage of mail surveys even if the questionnaires are delivered. By combining these two frames, we expect to reach a larger proportion of the population.

In addition to increased coverage, a multi-frame sampling design enables us to support the use of multiple modes of data collection, which may enhance response rates. We propose the use of CATI for samples drawn from an RDD frame, and web and mail surveys for samples drawn from an ABS frame. While it is possible to conduct multi-mode (telephone, mail, and web) with a sample from a single frame by offering a menu of mode options or choices, this approach has two major limitations. First, per frame (dualframe RDD or ABS), addresses can be “matched” to telephone numbers (and vice versa) for a limited proportion of the frame. In other words, telephone numbers are not available for all addresses in the ABS frame, and addresses are only available for a limited portion of the RDD frame. This inherently limits the ways in which we can contact a portion of the respondents, which could introduce response bias. Second, providing respondents with a choice of modes can decrease response rates. Providing respondents with a choice for mode can cause “paralysis” and anxiety, which can lead to inaction. Additionally, the appeal of any one mode option may be diminished by the introduction of another. We believe our strategy of tailoring mode options to the underlying sampling frame will lower data collection costs, reduce non-response error, and increase population coverage/representativeness.

Further, we will use a responsive design approach in which we evaluate design alternatives and develop strategies to maximize response or improve data quality. The principles of responsive survey design (RSD) can be applied during data collection to develop design modifications that address certain circumstances that may arise in the field. The fundamental technique of RSD is to develop a design framework (e.g. alternatives, experiments) based on the uncertainties in the data collection; develop and monitor key quality metrics based on survey data and paradata; and implement corrective interventions as necessary. We maintain tracking and monitoring systems which allow us to evaluate fielding progress across several metrics, and use these data to make adjustment to sampling design and/or protocols on a quarterly basis.

Optimal Dual-Frame Design

The American Association of Public Opinion Task Force distinguishes between two landline and cell phone dual-frame designs: dualframe with overlap versus dual-frame without overlap. Dual-frame without overlap means we interview cell-only respondents via the cell phone sample and screen out those who report the use of a landline. Representation of the landline population comes entirely from the landline sample. Dual-frame with overlap means we interview cell-only respondents via the cell phone sample as well as those who report the use of a landline. Representation of the landline population comes from combining the landline respondents from the cell phone sample with the respondents from the landline sample.

We plan to use a dual-frame with overlap since there are cost and quality benefits over the screening design. First, many dual-users report using their cell phones as their primary phone. Like cell-only, the “cell-mostly” population—those who use their cell as their primary phone—may be under-represented in landline samples, despite theoretically being covered by the

landline frame. Second, dual-users in a landline sample are different on many measures than dual-users in a cell phone sample. And finally, the net cost for conducting interviews with all cell users (even those with a landline) is not substantially more than interviewing cell-only users. Since a respondent's status as a cell-only or dual-user is not known in advance of speaking with him or her, we speak to about two dual-users for every three cell-only users. By conducting interviews with the dual-users we reach through the cell sample, the landline sample volume is reduced; usually, this is a cost benefit. This allocation is "optimal"; it is the most statistically efficient allocation of sample—no other allocation results in lower variance for the same cost, maximizing the effective sample size for a fixed cost.

Weighting of this survey will be performed based on designed and implemented weighting schemes that have been used on a large number of multi-frame samples and state-based health surveys including the National Adult Tobacco Survey.³ Weighting serves at least 3 important purposes as it corrects for: unequal probabilities of selection into the survey sample, thus possibly reducing bias associated with selection procedures, including the dual frame sample, differential non-response among elements of the survey population, reducing bias associated with non-response, and frame coverage differences relative to the target population. Weights will be appended to each survey record in the final data file.

We will be using design weights, frame integration, and raking ratio adjustment. Design weights will be computed as the inverse of the probability of selection of the phone number or address from the sampling frame. For landline and ABS sample, we randomly select a single person within the household to complete the survey. We adjust for the within-household selection with a weight equal to the number of household members eligible for the survey. For cell phone samples we assume the cell phone is a personal device and therefore the within-household weight is equal to one. In the landline samples, households are selected with probability proportional to their number of telephone numbers, we adjust for multiple phone lines by dividing the weights by the number of telephone lines. For cell phone samples and ABS, this adjustment is equal to one. Frame integration is conducted by integrating the three frames in two steps using traditional dual-frame methods.⁴ Raking ratio adjustment is conducted using an iterative ratio adjustment to adjust for non-response and non-coverage (of the non-telephone population). We will use the algorithm and methods used in 2012 for the demographic targets, the weight trimming, and the demographic imputations for BRFSS and previous published rake and trim algorithms.⁵ Missing values for the weighting variables will be imputed based on the following strategy. All imputation will be done separately by state. Age will be imputed as the mean value for each gender and race category. Marital status and educational attainment will be imputed based on a nearest neighbor hot-deck algorithm. The algorithm will impute marital status and educational attainment from the same respondent if both are missing. Race, age, and gender will be used to determine nearest neighbors.

4. Test of Procedures or Methods to be Undertaken

The development of this study and procedures was a collaborative effort among CDC, FDA, USDA, the ten FoodNet sites and ICF using established methods and procedures. Feedback from this group was used to refine questions, ensure accurate skip patterns, and estimate time for survey completion. Pre-testing of the questionnaire will be conducted with up to 200 persons prior to survey launch, including cognitive-, usability-, and pilot testing. Pre-testing data will be only used to inform the final questionnaire and will not be included in the analysis. In-person cognitive testing of the questionnaire will be conducted by ICF with cognitive interview in

English and Spanish to identify comprehension, retrieval, judgement and response problems. In adapting and testing of the questionnaire assessments will be conducted by ICF to assure the same response to each question regardless of mode and any needed revisions will be made. Usability testing and quality assurance detailed in the tables below. ICF will conduct CATI, web, and mail interviews and respondent debriefing per surveillance site. Paradata and respondent data collected will be reviewed by ICF and CDC and modification to the questionnaire or programming will be performed as needed. ICF will also assess item non-response to identify any problematic questions or survey sections that may contribute to respondent break-off.

Usability Testing of Survey

Usability Tests	
Purpose	To find and correct problems with the interactive survey experience
What will we test?	Paper, telephone, and screen-based survey versions
How do we test?	For self-administrated paper and web-based surveys, we will show participants a survey prototype and ask them to complete specific tasks while "thinking aloud". The investigator will note errors and watch for places in the interface where the participant hesitates, seems confused, or expresses frustration. For CATI surveys, interviewers will be paired with respondents, and observers will note errors, slow-downs, and difficult interactions involving both interviewers and respondents.

Quality Assurance by Mode and Task

Task	Detailed Subtasks
Questionnaire Translation	<ul style="list-style-type: none"> Review the final translated document to ensure that the translation flows and that there are no typographical errors
Beta Test Web and CATI Programs	<ul style="list-style-type: none"> Review frequencies of randomly generated data Independent skip-check program written (runs nightly throughout fielding) Manual test of program by programmers and project management staff CDC reviews electronic test versions of the programmed surveys for review CATI/web programs revised as needed based on CDC review; steps repeated until CDC approves electronic test version of survey program Use of detailed and comprehensive QC checklists
Interviewer Training	<ul style="list-style-type: none"> Dedicated team of FoodNet Population Survey interviewers General two-day training FoodNet Population Survey training attended by CDC staff and led by ICF project management
Sampling	<ul style="list-style-type: none"> CDC staff reviews and approves sampling plan Prior to pilot-test and each sample load, Senior Statistician, Project Manager, and Call Center Manager (for CATI) make independent sample estimates and reach consensus prior to load. Programmer loads sample and confirms loaded sample amounts with Senior Statistician, Project Manager, and Call Center Manager (for CATI) Project Manager monitors sample performance at minimum weekly throughout fielding, in consultation with Statistician and Call Center Manager (for CATI)
Data Collection	<p><i>Mail Data Collection</i></p> <ul style="list-style-type: none"> Standardize/correct sampled addresses using CASS-Certified software (Mailbox Toolkit from Satori Software) Review print-ready media for alignment, accurate text, correct position of address Check to ensure that addresses are facing address window and all items are inserted Order sufficient postage Verify that envelopes have postage and are sealed Verify that initial sample counts match counts of printed and posted material Seed ICF and CDC staff in mailing Record master ID from returned surveys <p><i>CATI Data Collection</i></p> <ul style="list-style-type: none"> CDC staff monitors interviews Data collection team monitors at least 10% of all interviews Professional project staff monitor interviews Monitor each interviewer at least once per week Assign supervisors to manage a team of no more than 10 interviewers Review call center shift reports and internal project tracking reports daily Skip-check program (checks survey paths as well as unexpected dispositions) runs nightly throughout fielding and the project team receives automatic e-mail notification of potential discrepancies

Task	Detailed Subtasks
	Web Data Collection <ul style="list-style-type: none"> Review frequencies from randomly generated data to ensure that the program is organizing data properly and recording values according to the survey specification Develop automated skip-check program to check data against defined conditions specified in the MS Word version of the questionnaire Nightly, check all data collected using automated program Hourly, check for and respond to Helpdesk e-mail Hourly, respond to calls and voice-mails received through the FoodNet IVR (Interactive Voice Response) system
Data File Preparation	<ul style="list-style-type: none"> Identify incomplete interviews and merge back into the main data file Clean and, when applicable, back-code open-ended responses Assign a final disposition to each record Produce frequency tabulations of every question and variable to detect missing data or errors in skip patterns
Weighting	<ul style="list-style-type: none"> Confirm that the sum of the base weights equals the frame counts in each stratum. Weight multipliers calculation confirmations and checks Dual-frame adjustment confirmations and checks Calibration (post-stratification/raking) confirmation and checks
Reporting	<ul style="list-style-type: none"> Manually review all numbers in report Senior Statistician review of inferences and recommendations Editor review to ensure no spelling errors, tables and exhibits numbered sequentially, narrative is clear and concise
Data Entry and Open-ended Responses	<ul style="list-style-type: none"> Range, consistency checks, and verification questions built into web and CATI programs, prompting respondents and interviewers to verify responses and reducing data entry error for numeric entries QA staff in the call center and Project Manager monitor and score interviewers, rating them on proper entry Project Manager reviews open-ended and "other/specify" responses after pre-test and biweekly throughout fielding to identify potential coding or training issues

Tracking and monitoring systems will evaluate progress across QA metrics listed above and these data will be used to make any needed adjustments to sampling design and/or protocols on a quarterly basis.

5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

This survey has been reviewed by analysts and statisticians within the Enteric Diseases Epidemiology Branch (EDEB) and the Division of Foodborne, Waterborne, and Environmental Diseases (DFWED). The contracting company, ICF International, has experience conducting dozens of multi-mode surveys over the past several years. ICF employs >200 survey and research professionals with advanced degrees and expertise in questionnaire design, sampling, data collection, weighting, data managements, analysis and has 25 years of continuous data collection for BRFSS and three previous cycles of the FoodNet Population Survey. Statisticians, PhD-level epidemiologists, and MPH-level analysts who work within EDEB will perform analysis of the final survey data.

References

1. ICF International. <http://www.icfi.com/services/survey-research>. Accessed February 5, 2016
2. CDC. Behavioral Risk Factor Surveillance System (BRFSS) Overview 2013. http://www.cdc.gov/brfss/annual_data/2013/pdf/overview_2013.pdf. Accessed April 21, 2015.

3. National Adult Tobacco Survey for 2013-2014. <https://govtribe.com/project/national-adult-tobacco-survey-for-2013-2014-and-2014-2015-cycles/activity>. Accessed February 5, 2016
4. Hartley H. Multiple Frame Surveys. Proceedings of the Social Statistics Section of the American Statistical Association. 1962;203-206.
5. Izrael D, Battaglia MP, Frankel MR. Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a. Raking). SAS Global Forum. 2009.
6. Hardnett FP, Hoekstra M, Kennedy M, Luenda, Angulo, F. Epidemiologic issues in study design and data analysis related to FoodNet activities. CID 2004; 38 (Suppl 3).

List of Attachments

Note: Attachment is included as separate file as instructed.

- **Attachment D: Pop Survey cleared questionnaire February 2016**